Authors' Response. Submitted to *Behavioral and Brain Sciences*, 21 June 2013

MIRROR NEURONS:

TESTS AND TESTABILITY

Caroline Catmur[1], Clare Press[2], Richard Cook[3], Geoffrey Bird[4], & Cecilia Heyes[5,6*]

[1]Department of Psychology, University of Surrey

[2]Department of Psychological Sciences, Birkbeck College, University of London

[3]Department of Psychology, City University London

[4]Social, Genetic and Developmental Psychiatry Centre (MRC), Institute of Psychiatry, King's College London

[5]All Souls College, University of Oxford

[6]Department of Experimental Psychology, University of Oxford

*Corresponding author: cecilia.heyes@all-souls.ox.ac.uk

Abstract

Commentators tended to focus on the conceptual framework of our article, the contrast between genetic and associative accounts of mirror neurons, and to challenge it with additional possibilities rather than empirical data. This made the empirically-focused comments especially valuable. The mirror neuron debate is replete with ideas; what it needs now are system-level theories and careful experiments – tests and testability.

R1. Introduction: Types of challenge

The target article ("Mirror neurons: from origin to function" – henceforth "O2F") called for a new approach to research on mirror neurons (MNs): a shift in priorities from 'ideas' to rigorous, theory-based experimentation. Reflecting current priorities, the commentators offered many more conceptual than empirical challenges; more in principle possibilities than testable hypotheses. In some respects this was disappointing, but it made the empirically-grounded commentaries especially precious, and, since many of the possibilities were very interesting, did not detract from the liveliness of the discussion.

We will respond to all substantive comments in roughly the order in which the relevant issues arose in O2F. The major exception is R2, which deals with comments relating to sections 2 and 7 of the target article. R3 relates to section 3, R4 to section 4, and R5 to section 5. Breaking the rhythm, R6 focusses on commentaries that offered convergent evidence in support of the associative account or extensions of our approach, and R7 relates to section 8 of O2F. Those who are quick on their feet will have noticed that there is no partner for section 6 of O2F, in which we presented experimental evidence in support of the associative account. That is because this evidence was not challenged. Some commentaries, discussed in R2, said that it does not show all that we think it shows, but no one denied that the properties of MNs can be changed radically by sensorimotor experience.

R2. Nature *and* nurture: Five false claims

Perhaps it is human nature to dichotomize nature and nurture (Linquist, Machery, Griffiths, & Stotz, 2011)! Maybe we humans have such a powerful tendency to split nature and nurture – to think of biological form as a product of internal *or* external factors, genes *or* environment, evolution *or*

learning - that we imagine others are doing it even when they are not. Such a cognitive bias would help to explain why many commentators mistook our genetic-associative distinction for a nature-nurture dichotomy. To put the record straight, here are five claims that we did *not* make in O2F because, like the commentators cited in this list, we think these claims are false: 1) Associative learning is sufficient for the development of MNs (**Bertenthal**; **Bonaiuto**; **Giese**; **Oberman, Hubbard, & McCleery**). 2) The genetic account assumes that learning does not contribute to the development of MNs (**Behme**; Bertenthal). 3) Genetically inherited traits are insensitive to experience (Bertenthal; **Kilner & Friston**; **Lingnau & Caramazza**; **Martin & Santos**). 4) Genetically inherited traits are insensitive to associative learning (Bertenthal; Bonaiuto; Kilner & Friston; Lingnau & Caramazza; **Newman**). 5) Associative learning cannot be adaptively specialized (Behme; Bertenthal; **Gervais**; Martin & Santos; **Simpson, Fox, Tramacere, & Ferrari**).

R2.1. Associative learning is sufficient for the development of MNs

Taken at face value, this claim is so obviously false that no one could imagine it was the thrust of our article. Even when one takes 'the development of MNs' to be shorthand for 'the development of the matching properties, or mirrorness, of MNs' – as we did throughout O2F – it is clear that the developmental system draws on a multitude of resources in addition to associative learning, starting with basic nutrients and the genetically inherited potential to develop neurons at all. The useful version of the 'associative learning is sufficient' claim was laid out by **Bonaiuto**, who suggested that the associative account assumes "every neuron is either directly or indirectly connected with every other neuron in the brain". This is useful because it gives us a new and clearer way of saying: "The associative hypothesis assumes that gene-based natural selection has played a significant background role with respect to the development of MNs; for example, in shaping the anatomy of visual and motor cortex for visual guidance of action" (O2F, sect. 2.2). With Bonaiuto's help, we can now rephrase this: The associative hypothesis does *not* assume complete or random connectivity between visual and motor cortex. Rather, it assumes that gene-based natural selection has

produced partial connectivity as an adaptation for the development of visual control of actions, not for mirroring. For example, visual areas detecting highly action-relevant properties of objects may have the potential to be more richly or readily connected with premotor cortex than visual areas processing less action-relevant properties of objects. It is also possible that genetic evolution has produced some preferential connectivity between visual areas processing observed action and motor areas. However, the associative account suggests that, if this is the case, these predispositions do not favor mirror over non-mirror or counter-mirror connections. Therefore, the issue that separates the associative account and Bonaiuto's view from the genetic account (e.g. **Fogassi**) is whether the pattern of partial connectivity, or "primary repertoire" (**Kilner & Friston**), was favored by natural selection because it tends to produce sensorimotor neurons that are good for guiding instrumental action, or because it produces mirror neurons that are good for action understanding or another social cognitive function.

R2.2. The genetic account assumes that learning does not contribute to the development of MNs

According to **Behme** (see also **Martin & Santos**)**,** our discussion of MN responses to unnatural stimuli – such as plastic crumpling and paper tearing – made clear that, although we said that "the genetic hypothesis does not necessarily assume that experience plays a minimal role" (O2F, sect. 3.1), the genetic account is really a straw man who insists that learning does not contribute to the development of MNs. By re-affirming its basic tenets, **Fogassi**'s commentary confirmed that the genetic account is not a straw man. More specifically, we agree with Behme that the mere fact that some MNs respond to unnatural stimuli is problematic only for the starkest version of the genetic account. However, as we emphasized (O2F, sect. 3.1), many MNs do not merely respond, they respond *maximally*, to unnatural stimuli. This is hard to reconcile even with versions of the genetic hypothesis that assign important roles to sensory and/or motor learning, because these assume that links among sensory neurons, or among motor neurons, account for responses to unnatural stimuli. In contrast, maximal responding to unnatural stimuli implies that the learning that produces tool-use

and audio-visual MNs is sensorimotor; it connects sensory neurons coding unnatural stimuli directly with motor neurons coding action (Cook, 2012).

R2.3. Genetically inherited traits are insensitive to experience

We are grateful to **Kilner and Friston** and **Lingnau and Caramazza** for mentioning some of the classic studies showing unequivocally that genetically inherited traits are sensitive to experience.

R2.4. Genetically inherited traits are insensitive to associative learning

The fourth claim deserves closer scrutiny. Several commentators assumed not only that we are making this claim, but also that, without it, the evidence that sensorimotor learning can radically change MNs does not support the associative over the genetic account (**Bertenthal**; **Bonaiuto**; **Gervais**; **Kilner & Friston**; **Newman**). This misconception is largely our fault. The argument underlying our tests of the associative account has been explained previously (Heyes, 2010), but was not made explicit in the target article. To rectify this, here is a short, focused version of the argument: Our training experiments (reviewed in O2F, sect. 6) indicate that (A) counter-mirror sensorimotor experience (e.g. moving an index finger whenever you see little finger movement) reverses the matching properties of MNs, and (B) sensorimotor experience with inanimate objects (e.g. making a fist whenever you see a certain shape) induces MNs to respond to inanimate objects[1]. As predicted by the associative account, effects A and B show that sensorimotor experience perturbs but does not damage MNs: it prevents MNs from selectively encoding similar observed and executed actions, but does not stop them from encoding systematic relations among stimuli and responses. This is hard for the genetic account to accommodate because 1) if a trait is a gene-based adaptation, its development tends not to be perturbed by environmental variations that were present when the trait evolved, and 2) sensorimotor experience of the kind that reverses and induces MN activity is likely to have been present in the period when, according to the genetic account, MNs evolved. Specifically, it is likely that the common ancestors of extant monkeys and humans experienced

contingencies between observation and execution of non-matching actions (e.g. when one individual countered a blow from another), and between objects and actions (e.g. when distinctive actions were made on distinctive objects). Thus, our key assumption is not that 'genetically inherited traits are insensitive to associative learning', but that 'if a trait is a gene-based adaptation, its development tends not to be perturbed by environmental variations that were present when the trait evolved'[2].

R2.5. Associative learning cannot be adaptively specialized

Several commentators suggested that we were wrong to contrast the genetic account with the associative account because associative learning can be adaptively specialized (aka 'canalized' and 'exapted') (**Behme**; **Bertenthal**; **Gervais**; **Giese**; **Keysers, Perrett, & Gazzola**; **Lotem & Kolodny**; **Martin & Santos**; **Simpson et al.**). For example, as indicated by Garcia's classic experiments on food aversion learning (e.g. Garcia, Kimeldorf, & Koelling, 1955), genetic evolution can alter input mechanisms or learning rate parameters so that contingencies between some pairs of events (e.g. flavor and illness) are learned more readily than others (e.g. flavor and shock); a phenomenon known as 'selective association' (Gemberling & Domjan, 1982; Heyes, 2003; Pearce, 2008). To be clear: we have no doubt that associative learning can be adaptively specialized. Indeed, if we had any doubts, we would not have put so much effort into experiments that were explicitly designed to test for signs of adaptive specialization in the development of MNs. We have tested for adaptive specialization - e.g. using counter-mirror and induction training (O2F, sect. 6) - and failed to find it. That is why the associative account is not a hybrid model; why it is "extreme" (Lotem & Kolodny) in suggesting that MNs are forged by domain-general processes of learning that have not been specialized by genetic evolution for the development of MNs.

The associative hypothesis would be less plausible if it had been shown that adaptive specialization of associative learning is, not merely possible, but normal or typical (**Lotem & Kolodny**). However,

the evidence does not support typicality. There are quite a few examples of selective association, but selective association can be due to prior learning (Mackintosh, 1973) rather than genetic processes – 'ontogenetic inflection' rather than 'phylogenetic inflection' (Heyes, 2003) – and in many cases it has merely been assumed that the source of selective association is phylogenetic; that it constitutes adaptive specialization.

The commentary articles offered a rich array of hybrid genetic-associative possibilities. For example, they pointed out that MNs could be built by a form of associative learning that is adaptively specialized for empathy (**Gervais**) or self-monitoring (**Bonaiuto**) rather than action understanding; via epigenetic as well as genetic processes (**Simpson et al.**); by the modification of memory parameters as well as input mechanisms (Bonaiuto; **Keysers et al.**; **Lotem & Kolodny**); and where the modified input mechanisms involve mother-infant face-to-face interaction, infant-directed speech, imitation by adults, and the face preference (Gervais; Keysers et al.). We agree that these are all logical possibilities, and that there is "no *a priori* reason" (**Behme**), such as lack of parsimony (**Lotem & Kolodny**), to reject hybrid models. Indeed, as we said in the target article, we see hybrid modeling as "a promising direction for future research" (O2F, sect. 7.2). We just want to see some evidence. It is not enough to generate hypotheses; they also need to be tested.

Reader's commentary confirms that it is possible to test adaptive specialization hypotheses by discussing research on social learning, rather than MNs, in which this has been done successfully. In contrast, very few commentators identified empirical research on MNs that, in their view, supports a hybrid, adaptive specialization model over the associative and genetic accounts. **Oberman et al**. cited a study by Howlin and Moss (2012) which reported that people with autism spectrum conditions (ASC) are at a disadvantage, compared with neurotypical controls, in relation to employment, social relationships, health, and quality of life. To count this as evidence for a hybrid model, it would be necessary to make three unjustified assumptions: that people with ASC have

abnormal MNs; that their disadvantages are due to this abnormality; and that the disadvantages would not occur if MNs developed in the manner described by either the genetic or the associative account.

Similarly, citing Meltzoff (1995), **Bertenthal** suggested that infants more readily imitate human than mechanical movements, and that this animacy bias supports a hybrid model over the associative account. However, a replication of Meltzoff's (1995) study found evidence that his results were due to stimulus enhancement – a phenomenon that does not implicate MNs – rather than imitation (Huang, Heyes, & Charman, 2002). More broadly, animacy bias in imitation – which has been reported in other studies – is fully compatible with the associative account. Human developmental environments typically offer many more opportunities to associate the observation of human action, than the observation of mechanical action, with action execution (Press, 2011). Therefore, one does not need to invoke adaptive specialization to explain animacy bias. Furthermore, as predicted by the unadorned associative account, animacy bias is abolished when action execution is correlated with observation of mechanical movements (Press, Gillmeister, & Heyes, 2007).

Taking a broader approach, **Gervais** (see also **Martin & Santos**, and **Oberman et al.**) suggested that hybrid models are better able to explain why monkeys and humans, but not rats, have MNs. The first problem with this argument is that we do not know whether rats or free-living monkeys have MNs; they have not been tested. The second problem is that the associative account predicts species differences based on the amount and kind of sensorimotor experience received in the course of typical development, and – borrowing **Bonaiuto**'s terminology again – the pattern of partial connectivity between sensory and motor areas that has been favored by genetic evolution for visual guidance of action.

We hope our article will encourage those who back hybrid models to test them against the genetic and associative accounts in carefully designed experiments. This was, to a large degree, our purpose in writing O2F. The resulting experiments may well show that the associative account is wrong or needs to be augmented (Heyes, under review). For example, in common with **Gervais**, we have suggested elsewhere that cultural evolution, rather than genetic evolution, may have favored social practices that facilitate the development of imitation and MNs through associative learning (Heyes, 2013). But even if future experimental research shows that the associative account is plain wrong, in our view it would be a good result. It would mean that research on MNs has moved from exciting speculation to exciting experimentation; from ideas to tests and testability.

R3. 'Goals' and the field properties of MNs

In section 3 of O2F, we argued that research on the field properties of MNs does not support the claim that they selectively encode action 'goals'; a central plank of the case for the genetic account. Several commentators explicitly endorsed our view that MNs do not selectively encode goals (**Giese**; **Heil, van Pelt, Kwisthout, van Rooij, & Bekkering**; **Kilner & Friston**). Others repeated the claim that MNs selectively encode goals, but did not engage with the evidence we presented (**Brown & Brüne**; **Fogassi**; **Gallese & Sinigaglia**; **Khalil**; **Martin & Santos**). They did not explain how the goal-encoding claim can be reconciled with the evidence that MNs are responsive to intransitive as well as transitive actions, and to low-level features of action such as the path through space, effector used, and observer's viewpoint.

Orban mentioned three studies that we did not discuss in O2F, two using fMRI and one using single unit recording in monkeys. The first fMRI study (Nelissen, Luppino, Vanduffel, Rizzolatti, & Orban, 2005) reported that area F5c was active during action observation when the actor's body was in full

view but not when the monkey could see only parts of the actor's body. This finding does not warrant the general claim that a full view of the agent is necessary for MN firing in monkeys. For example, audiovisual MNs discharge when the monkey hears paper tearing and can see only the experimenter's static, disembodied head and arms (see Figure 1B, Keysers et al., 2003). However, in those cases where MN activity is sensitive to the visibility of the actor's body, the associative account makes the testable prediction that this is due to either generalization decrement or contextual learning of the kind discussed in section 4 of O2F. The second fMRI study (Nelissen et al., 2011) confirmed that signals travel from STS to F5c over two parietal way stations, which is entirely compatible with our statement that some sensory neurons in STS are "connected, directly or indirectly, to motor neurons in PMC … and parietal cortex" (sect. 2.2, O2F).

The third study reported 'suppression MNs'; neurons that discharge when a focal behavior is executed and show below baseline activity when a similar behavior is observed (Kraskov, Dancause, Quallo, Shepherd, & Lemon, 2009; Vigneswaran, Philipp, Lemon, & Kraskov, 2013). This study is very interesting indeed, but it is not clear why **Orban** believes that it supports a genetic or hybrid account over the associative hypothesis. Indeed, at first blush, suppression MNs seem to be troublesome for any account suggesting that evolution designed MNs to promote action understanding. If standard 'excitation MNs' were designed to promote action understanding, it looks as if suppression MNs must have been designed to prevent action understanding. This is implausible, and therefore supporters of a genetic or hybrid account have appealed to an untested auxiliary hypothesis – that the specific and presumably unique function of suppression MNs is to allow self-and other-produced actions to be distinguished (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010).

Given that associative learning can be both excitatory and inhibitory, there is no reason to doubt that suppression MNs are compatible with the associative account. However, a good deal more will need to be discovered about suppression MNs before the associative account, or any other

hypothesis, can offer a clear and specific explanation of their properties. For example, we need to know how suppression MNs are related to MNs that show below baseline activity during *both* action execution and observation (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992); about the source(s) of inhibitory input (frontal executive regions, parietal cortex, and/or other F5 neurons), and about the developmental history (e.g. the training received by the animals prior to recording) of the monkeys in which suppression MNs can be found.

Fogassi suggested that two observations favor the genetic account, which postulates goal encoding, over the associative account: the activity of some MNs covaries with the outcome of the observed movement (grasp-to-eat versus grasp-to-place), and "the percentage of MNs tuned for the hard-wired action (grasp-to-eat) is much higher than that of MNs tuned for the learned action (grasp-to-place)". We discussed the first of these observations in O2F (sect. 4.2), showing that the pattern of covariation can be explained by contextual associative learning. We cannot see how the second observation bears on the distinction between the genetic and associative accounts as they were described in the target article, or why the commentator identifies grasp-to-eat as "hard-wired" and grasp-to-place as "learned", but it should in any case be clear that the associative account predicts varying proportions of MNs as a function of the differential contextual sensorimotor experience received by the animal.

R4. Still the right kind of learning

In section 4 of O2F we argued that domain-general principles of associative learning, including contingency sensitivity and contextual modulation, provide a unified explanatory framework for the field properties of MNs. While many commentators agreed that associative learning plays a significant role, some argued that a special kind of associative learning is involved in the development of MNs. In R2 we responded to commentaries suggesting that it is special in being

adaptively specialized. In this section we turn to commentaries suggesting that the development of MNs depends on associative learning that is more "Hebbian", more "ideomotor", and more "high-level" than that which has been revealed by research on conditioning in animals.

R4.1 Hebbian and/or associative?

In O2F we offered two interpretations of what **Keysers et al**. described as the Hebbian hypothesis: on one reading, it is "identical in substance to the associative hypothesis [and] helpful in providing a more detailed neuronal model of how sensorimotor experience makes MNs out of motor neurons", and on the other reading, which we thought "less likely to represent the authors' intentions, 'Hebbian learning' differs from 'associative learning' in depending on contiguity alone, rather than both contiguity and contingency" (sect. 7.1). In their commentary, Keysers et al. seemed to confirm that our first interpretation was correct. They certainly emphasized the neuronal specificity of their model, and stressed that, however Hebbian learning may have been understood in the past, Keysers et al. understand it to be a process that depends on contingency. However, if, as they suggested, the associative and Hebbian accounts are merely "descriptions at different levels", and it is "idle" to argue that one is better than the other, it is not clear why Keysers et al. also suggested that the two accounts make different predictions; that the Hebbian account would, and the associative account would not, predict the formation of associations between as well as within action phases - for example, between "reach-STS" (observation of reaching) and "grasp-PM" (execution of grasping), as well as between "grasp-STS" (observation of grasping) and "grasp-PM" (execution of grasping). Two theories are either equivalent, and it is idle to ask which is correct, or they make different empirical predictions, and it is the basic stuff of science to test them against one another.

In this case, we think that **Keysers et al**. were right about the equivalence rather than the differential predictions. Tracking of predictive relationships between events (rather than event co-occurrence), via prediction error, is at the heart of the formal models of learning that underwrite the associative

account (e.g. Rescorla & Wagner, 1972; Schultz & Dickinson, 2000). Therefore, the associative account is wholly compatible with "sensory prediction" (Keysers et al.), "predictive coding" (**Brass & Muhle-Karbe**; **Kilner & Friston**; Press, Heyes, & Kilner, 2011), and, most specifically, with the formation of associations between, as well as within, action phases. The challenge for both the associative and the Hebbian views – assuming they are equivalent in substance – is to define the conditions in which learning produces 'classical' MNs, which fire during the observation and execution of the same action (e.g. grasping), and those in which it yields non-matching MNs such as logically-related MNs, that may encode relations between successive components of an action sequence. The advantage of the associative / Hebbian account over the genetic account is that it can readily accommodate both kinds of MNs because associative learning is all about prediction rather than matching. However, since associative learning does prediction on a variety of timescales, those who support an associative account need to work on predicting the timescale on which it will operate for different types of action and of sensorimotor experience.

R4.2 Ideomotor and/or associative

In their thoughtful commentary, **Brass and Muhle-Karbe** suggested that MNs are a neural substrate of ideomotor representations, established through the formation of response-effect associations, and contrast this with our associative account, which identifies MNs as "sensorimotor" representations, and refers to "sensorimotor" learning (see also **Ho, MacDonald, & Swain**). They may well be right. There may be a real, empirically testable difference between their ideomotor account and our associative model. But, as with **Keysers et al.**, where Brass and Muhle-Karbe see conflict, we see harmony. We suspect that the two views *seem* to be different because what we call "stimuli" they call "response effects", but the two accounts are in fact united in suggesting that the critical associations – the associations that give MNs their cardinal properties – link neuronal representations of "R" and "E", not "S" and "R", in their diagram (Brass & Muhle-Karbe, Figure 1). If this is correct, our training experiments (sect. 6.1, O2F) contain the right ingredients for the

ideomotor account – they expose participants to contingent relationships between R and E, execution and observation of action – but not in the order in which these events are typically experienced in everyday life. In each trial of our training regimes, participants observe an action and then execute an action. We have used this kind of regime because it is easier to control the contingencies experienced by participants when the observed action comes first, and because there is compelling evidence that associative learning is bidirectional (Ayres, Haddad, & Albert, 1987; Elsner & Hommel, 2001, 2004; Mahoney & Ayres, 1976). It is important to note that, confirming the bidirectionality assumption, other groups have replicated and extended our findings using training regimes in which response execution precedes action observation (e.g. Wiggett, Hudson, Tipper, & Downing, 2011).

Of course, the proof of the ideomotor / associative pudding will be in the testing, but the examples mentioned by **Brass and Muhle-Karbe** do not yet convince us that it will be possible to generate differential predictions. They suggested that two empirical findings are inconsistent with the associative account. In the first, Kunde (2001) found that responses were executed faster when followed by compatible, rather than then incompatible, sensory effects; evidence for anticipatory excitation of the sensory consequences of action. However, as discussed in the previous paragraph, the associative account is compatible with predictive motor-to-sensory propagation, following exposure to either S-R or R-E contingencies. In the second study, Liepelt, von Cramon, and Brass (2008) found a larger index/middle automatic imitation effect when fractional finger movements were presented inside, rather than alongside, metal clamps. Applying the associative account, Heyes (2011) agreed with Brass and Muhle-Karbe that this could be described as an effect on mirroring of "high-level beliefs about the intentionality of the observed action", but suggested that the effect was mediated by enhanced attention to the clamped movements (see also **Heil et al**., and **Brown & Brüne**).

R4.3 Reasoning and/or association

The associative account identifies the learning processes involved in the development of MNs with those that produce conditioning in a variety of animal species. However, **Heil et al.** argued that this type of 'low-level' associative process is ill-suited to producing MNs. Appealing to evidence that beliefs about causal structure can modulate human associative learning, Heil et al. raised the possibility that another type of associative process is responsible for the development of MNs; one that is shaped by and integrates higher-level cognitive processes. This is an interesting but, in its current form, largely speculative hypothesis. To make it testable, **Heil et al.** would need to say more about the kind of high-level processes they take to be involved in the development and on-line control of MNs (e.g. the processes that yield contingency awareness may or may not be reasoning processes), and about the way in which the high- and low-level processes are supposed to interact. In humans, language-mediated high-level processes can certainly influence associative learning by modulating input mechanisms - for example, task instructions can direct participants' attention towards or away from the computer screen on which stimuli are presented – but this kind of interaction is entirely compatible with our associative hypothesis (e.g. **Lotem & Kolodny**). To motivate their hypothesis, **Heil et al.** would also need to cite findings from the MN literature which they regard as consistent with a high-level associative process, but inconsistent with a low-level mechanism. At present they suggest only that a high-level process would be needed if MNs are to mediate action understanding. This may be so, but it is not a compelling argument for a high-level process given the ambiguity of the term 'action understanding' (sect. 2.1, O2F), and the corresponding lack of clear evidence that action understanding depends on MNs (sect. 8.2, O2F).

In R4 we have argued that both the associative and Hebbian accounts are compatible with predictive coding; that the associative account is complementary to, not in conflict with, ideomotor theory; and that there are currently no data from MN research suggesting that their development or operation is guided by a 'higher-level' associative process. We therefore maintain that phylogenetically ancient

associative processes are 'the right kind of learning' to yield the field properties of MNs.

R5. Poverty of the poverty argument

In section 5 of O2F we argued that human developmental environments typically contain multiple sources of the kind of correlated sensorimotor experience necessary to build MNs; each of these sources is rich; and the mechanisms of associative learning can make swift and efficient use of these sources. No commentary queried the evidence in support of this 'wealth of the stimulus' argument. We also suggested that imitation and EEG data from human and monkey neonates do not support a complementary 'poverty of the stimulus' argument; they do not show that MNs develop too soon to be built by associative learning. Some commentators concurred with this view. For example, **Oberman et al.** agreed that the evidence for neonatal imitation is limited, and suggested that, even if the evidence were stronger, it is unlikely that neonatal imitation would reflect MN activity because the cortical areas containing MNs are not fully developed at birth. Similarly, **Rizzolatti** found it hard to "conceive how the mirror mechanism of a neonate might have a neurological maturity such as to provide a precise copy of tongue protrusion". However, two commentaries opposed our view of the neonatal imitation literature: **Gallese and Sinigaglia** did so implicitly, by simply stating that neonates imitate, but **Simpson et al.** cited evidence which, in their view, shows that neonatal imitation is a "real phenomenon".

There are, in our view, several problems with **Simpson et al.**'s survey of the literature on neonatal imitation. First, in pointing out that neonatal imitation has been demonstrated "in over two-dozen studies", it focused on positive effects, and disregarded the greater number of published studies reporting negative effects; failure to find evidence of neonatal imitation (see Figure 2, in O2F). Second, the vast majority of the positive findings cited by Simpson et al. came from studies that incurred a high risk of Type I error because they did not use the standard, 'cross-target' or 'double-

dissociation' procedure. This procedure reduces the risk that non-specific effects of a movement stimulus will be mistaken for imitation (production of a specific, matching response) by comparing the frequency of action A following observation of A and B, *and* the frequency of B following observation of A and B, i.e. looking for an interaction between stimulus and response. The only evidence of such an interaction cited by Simpson et al. (Ferrari et al., 2006) could have occurred by chance; it appeared to be present on only one of four test days and across only the lipsmacking-tongue protrusion comparison. Consistent with this possibility, the effect appears not to have been replicated in a subsequent study by the same authors (Paukner, Ferrari, & Suomi, 2011; see O2F for discussion). Finally, Simpson et al. did not offer compelling reasons to believe that the mixture of positive and negative findings, illustrated in Figure 2 of O2F, is due to individual differences in neonatal imitation rather than a general lack of imitative capacity in newborns. Yes, it is possible that neonates vary in their ability to imitate, and they are certainly difficult to test. But in humans the tongue protrusion effect has been replicated many times, suggesting that current procedures are good enough to detect a signal when a signal is there to be found.

We argued in O2F that the tongue protrusion signal is reliable but indicative of an oral exploratory response rather than imitation (Jones, 1996, 2006). Contesting this view, **Simpson et al.** cited a new study (Nagy, Pilling, Orvos, & Molnar, 2012) indicating that observation of tongue protrusion does not induce arm and leg movement. Appendage movements are unlikely to be part of oral exploration, and therefore this study does not bear on our point.

Turning from imitation to EEG measures, **Simpson et al.** and **Oberman et al.** argued on two grounds that alpha suppression during action observation is a valid indicator of MN activity in neonatal monkeys: 1) when human adults observe action, alpha suppression has been found alongside BOLD responses in premotor cortex, and 2) some studies have found that when people with ASC observe action, they show less alpha suppression than neurotypical controls. The first of these observations

does not establish alpha suppression as a valid index of motor, rather than somatosensory activity, because alpha effects correlate with both premotor and somatosensory BOLD responses (Arnstein, Cui, Keysers, Maurits, & Gazzola, 2011); MEPs during action execution correlate with beta, but not with alpha, suppression (Lepage, Saint-Amour, & Théoret, 2008); and in adults the source of alpha suppression during action execution has been traced to somatosensory cortex (Hari & Salmelin, 1997). It is likely that, due to high connectivity between somatosensory and motor areas, motor activity sometimes contributes to alpha suppression. However, even if a motor contribution could be demonstrated in studies of action observation in infants, it would remain an open question whether the motor contribution represented mirroring – activation of motor programs corresponding to the actions observed – or a generalized readiness to act. The force of the second observation – that people with autism show less alpha suppression than controls – rests on the highly controversial assumption that people with autism have abnormal MNs (Southgate & Hamilton, 2008; Bird, Leighton, Press, & Heyes, 2007; Press, Richardson, & Bird, 2010).

Research on mirroring in neonates faces formidable methodological challenges. Unless or until these are overcome, it cannot support a compelling poverty of the stimulus argument.

R6. Convergent evidence and extensions of the associative account

Many commentators presented convergent evidence supporting the associative account. **Oosterhof et al.** reported multivariate pattern analyses (MVPA) of fMRI data that indicate mirror – sensorimotor matching – responses in brain areas outside the classical parietal-frontal mirror circuit. As Oosterhof et al. note, the finding that mirror responses are widespread across the brain is in line with the predictions of the associative account. It also therefore counters those commentators who argued that the associative account must be wrong because MNs are only found in restricted brain areas (**Bonaiuto**; **Fogassi**; **Martin & Santos**; **Orban**).

Saygin and Dick discussed evidence indicating that mirror responses are not solely responsive to biological stimuli, being present also for point-light displays and robotic or android movements. As they note, and particularly in the light of evidence that some MNs respond maximally to unnatural stimuli (see R2.2, and Cook, 2012), such findings fit more closely with an associative than with a genetic account of MN matching properties.

Presenting further convergent evidence for the associative account, several commentators mentioned data demonstrating experiential effects, both on social behavior more generally, and on mirror responses in particular. **Reader** outlined studies showing that associative learning allows both birds and insects to learn from conspecifics, while **Newman** discussed how experience modulates squirrel monkeys' vocal responses to other monkeys' calls. **Holt and Lotto** reported data demonstrating that in infants, cortical responses to perceived speech sounds in motor areas show a developmental profile consistent with experience-dependent learning; and **Saygin and Dick** discussed a study showing effects of sensorimotor experience on responses to perceived sounds in premotor areas (Dick, Lee, Nusbaum & Price, 2011).

A number of commentators suggested ways in which the associative account could be extended to reward processing, speech processing, and high-level social interaction. **Brown and Brüne** (see also **Giese**) suggested the associative account may provide a useful framework for conceptualizing the role of reward in the development and function of the mirror mechanism. This line of enquiry is well worth pursuing, but interpreting the role of reward in the operation of a mature mirror mechanism is unlikely to be easy. Whereas Brown and Brüne suggested that putative examples of contextual modulation may in fact reflect differences in the processing or value of reward, an alternative interpretation is that instances of putative modulation by reward reflect stimulus-driven contextual

control. Moreover, where participants attend closely to actions associated with reward or punishment, differential mirror responses may also reflect input modulation (Heyes, 2011).

De la Rosa and Bülthoff highlighted the value of contingency sensitivity and contextual modulation for understanding how humans behave in social interactions. We certainly echo this view. Contingency sensitivity offers an elegant account of the emergence of automatic imitation of some actions, but pre-potent complementary responses for others (Tiedens & Fragale, 2003). Moreover, interactive behavior is frequently ambiguous: the appropriate response to an observed action may be imitative in one situation, but complementary in another. Contextual modulation by physical features of the context (e.g. visual, auditory cues) or internal states (e.g. level of arousal, presence of hormones) may serve to resolve this ambiguity and thereby enable flexibly adaptive behavior (Cook, Dickinson, & Heyes, 2012).

Another commentary which addressed behavior in social interactions was that of **Duran, Dale, and Richardson**. These authors endorsed the associative account, but we disagree with their characterization of our account as "negative eliminativist": on the contrary, we believe our proposal for a new approach to investigation of MN function is a positive message, and it was viewed as such by most commentators (see R7.2). However, we applaud the ambition of their thesis, which seeks to use domain-general processes, including associative learning, to explain social interaction in naturalistic settings.

R7. Looking ahead

The final section of O2F called for a new approach to research on MNs, with a keen eye on participants' developmental history, and using system-level theory to define putative functions of

MNs, and rigorous experimental methods to test the system-level hypotheses. In this section we respond to comments on the theoretical (R7.1) and empirical (R7.2) components of these proposals.

R7.1 Understanding "action understanding"

In O2F we pointed out that there is "no consensus about exactly what is meant by 'action understanding', or how it differs from cognate functions such as 'action perception', 'action recognition', and 'action selection'" (sect. 2.1). This is a conceptual problem that has generated a major empirical problem. When it is not clear what a mechanism is supposed to be doing, it is all-but impossible to design experiments finding out whether the supposition is correct. We also suggested that the best way to solve this problem would be to focus on the development of system-level theories (sect. 8.1.2), in which MNs feature as one component of a system defined by its outcome or typical effect, and in which the role of MNs is clearly distinguished from, and related to, the roles of other components (Heyes, under review).

In the commentaries, no one denied that understanding "action understanding" is a problem, and many endorsed the need for system-level theory, explicitly or by discussing particular system-level theories with enthusiasm (**Behme, Heil et al., Holt & Lotto, Kilner & Friston, Lingnau & Caramazza, Saygin & Dick**). (We particularly liked Kilner and Friston's slogan, "no neuron is an island".) However, confirming that we had not constructed a straw man (**Oberman et al.**), members of the Parma group eschewed the system-level approach, seeking to elucidate the "action understanding" function of MNs by 1) repeating the claim that they encode or interpret action "goals" (**Fogassi; Gallese & Sinigaglia**), and are therefore "bearers of intentionality" (Gallese & Sinigaglia); 2) noting that MNs transform "sensory information into a motor format" (**Rizzolatti**); and 3) describing MN-mediated action understanding as "action understanding 'from inside' (Rizzolatti & Sinigaglia, 2010): a *first person process*, where the self feels like an *actor*, rather than a *spectator*" (Rizzolatti). The first of these strategies is not promising because, as we argued in O2F (sect. 3), the term "goal" is at

least as ambiguous as “action understanding”, and on at least two natural interpretations of “goal” – as an object of action, and as a high-level intention – the evidence suggests that goals are not selectively encoded by MNs. The second strategy is more straightforward, but it is clearly insufficient because it does not refer to the cardinal, matching properties of MNs. They can certainly be said to transform sensory information into a motor format, but that is also true of, for example, canonical neurons. The third strategy is phenomenological; it distinguishes MN-mediated action understanding from other sorts of action understanding with reference to the actor’s conscious experience – whether he or she “feels” like an actor or a spectator. This suggestion is coherent and intuitively appealing, but it is unlikely to be helpful as a guide for experimental research. Without a return to the methods of introspectionism, researchers in psychology and neuroscience are “spectators”; we are ‘on the outside’, trying to work out the function of MNs, not by examination of our own conscious experience, but by studying the brains and behavior of other people. Therefore, to use the phenomenological strategy we would need to be told how, from the outside, to distinguish ‘inside’ and ‘outside’ action understanding; how these types of action understanding differ in terms of the observable behavior they produce.

Rizzolatti also explained with helpful clarity that he now believes that MNs contribute to social cognitive functions such as imitation, empathy and song recognition, as well as action understanding, and regards “the problem of how mirror neurons originate [as] utterly irrelevant as far as their function is concerned”. We have no quarrel whatever with these views; they are wholly in accord with the “functionally permissive” character of the associative hypothesis (O2F, sect. 8), and the way in which it splits questions about the origin of MNs from questions about their function. However, it is important to note that these views do not make the genetic hypothesis a straw man – **Fogassi**’s commentary, and the quotations in O2F (sect. 2.1), attest to its reality – or obviate the need for system-level theory. Indeed, they increase that need. There are just two widely recognized methods of defining the function of a trait in a biological system: one does it with reference to the

trait's history, typically its evolutionary history, and the other, 'functional analysis', does it in the context of what we have described as system-level theory (Cummins, 1975; Godfrey-Smith, 1994). Given that these are the two paths available, rejection of the history route – denial that the function of MNs relates to their origins – makes pursuit of the alternative, system-level theory, absolutely essential; and multiplying the putative functions of MNs means that we need system-level theories defining the role of MNs, not only in action understanding, but also in imitation, empathy and song recognition.

Gallese and Sinigaglia are absolutely right: the mere fact that "action understanding" is a new term is not a reason to reject it. All we are saying, here and in O2F, is that those who use the term need to tell us what it means, and to do so in the context of system-level theory.

R7.2 Experimental approaches

Few commentators (except perhaps **Fogassi**) disagreed with our proposal that MN research should report participants' developmental history: that is, their sensory, motor, and sensorimotor experience with the actions for which MNs are being tested. Indeed, **Gervais** commented on the "value of the enterprise" of documenting the effects that variation in developmental environment may have on the ontogeny and function of MNs. In this vein, the developmental data described by **Krogh-Jespersen, Filippi, and Woodward** are a promising start. These data demonstrated how infants' motor experience (specifically, in this case, their tendency to produce goal-directed actions) relates to their neural responses to observation of others' actions; to their ability to understand others' actions; and to their ability to imitate. These findings are consistent with the predictions of the associative account; however, they serve to highlight the importance of assessing participants' previous sensorimotor (not just motor; cf. **Gallese & Sinigaglia, Rizzolatti**) experience with specific actions when investigating neural responses to, understanding of, or imitation of, those actions (see earlier response to **Bertenthal** for further explication of this point).

Several other commentators appeared to embrace our suggestion that MN research pay closer attention to developmental history: **Simpson et al.** proposed examining the effects of early sensorimotor experience on imitation, while **Ho et al.** noted that it would be informative to consider the development of MNs in infants who have atypical experience, for example, those who cannot walk themselves or whose parents suffer from depression. We wholeheartedly concur that research with these populations could prove highly informative in dissociating the relative roles played by sensory, motor, and correlated sensorimotor, experience in the development of MNs, and perhaps also shed some light on their function.

We are pleased that most commentators endorsed our call for a more rigorous, experimental approach to the investigation of MN function (**Behme, Gervais, Holt & Lotto, Krogh-Jespersen et al., Lingnau & Caramazza, Newman, Oosterhof et al., Reader, de la Rosa & Bülthoff**). Here we address some criticisms of our proposals, and discuss the commentators' additional suggestions for future research.

Orban's commentary illustrated the constraints that have been placed on MN research to date by the genetic account. It suggested that MNs have only, and will only, be found for grasping actions; that MNs can only be considered such when found in classical areas; and that only a primate model holds any value for understanding the human brain. Such suggestions are in accord with an account that considers MNs to be a genetic adaptation for representing others' grasping actions in one's own motor system (see also **Fogassi**). They illustrate the risk that the genetic account will confine future research to certain actions, brain areas and species. However, as we indicated in the target article, the associative account predicts (and data support these predictions) that mirror neurons will be found for a range of actions (e.g. Ferrari, Gallese, Rizzolatti, & Fogassi, 2003), in a range of brain areas (e.g. Dushanova & Donoghue, 2010; Mukamel et al., 2010; Tkach, Reimer, & Hatsopoulos,

2007), and across a range of species (e.g. Prather, Peters, Nowicki, & Mooney, 2008). We therefore see no reason to reject the possibility that MNs could be induced in other species through sensorimotor training of species-appropriate actions.

We are not aware of any research that has investigated whether MNs are present in rodents. However, those commentators who denied that MNs are (**Gervais**), or could be (**Orban**), present in rodents (or in other species; **Oberman et al.**) seem to have misunderstood our point. The associative account suggests that whether MNs are likely to exist in laboratory (or indeed free-living) rats will depend on their prior learning history; but it makes the strong prediction that regardless of whether they are present or not, appropriate experience in which the sight or sound of an action is paired contingently with execution of that action should produce visuo- or auditory-motor MNs for that action, in rodents and indeed in any other species (an approach also endorsed by **Reader**'s commentary, in which he proposed the use of insect or fish models to investigate MNs). Building on the previous point, **Newman**'s commentary provided an interesting description of behaviors consistent with the existence of audio-vocal mirror-like circuits in a variety of species. We agree with Newman that it would be interesting to test for MNs in these species – indeed the presence of audio-vocal MNs in swamp sparrows has already been demonstrated (Prather et al., 2008).

The preceding commentators focused on experimentation into MN properties. Other commentators proposed investigation of the possible role of MNs in contagious behavior (**Provine**) and in higher-level social cognition tasks: **Martin and Santos** advocated carrying out single-unit recording while macaques undergo tests assessing their awareness of others' knowledge. We believe that such an endeavor would benefit from the kind of system-level analysis which we outlined in the target article (see also **Holt & Lotto**). Such an analysis would specify, for example, the role that MNs are predicted to play in these tests and whether or not this role must be carried out by MNs.

Several commentaries touched on the methodological challenges facing future MN research. **Lingnau and Caramazza** understood us to be suggesting that research into the modulation of MN responses by associative learning is sufficient to provide evidence of the role of MNs in social cognition. This is certainly not the case. Instead, the associative account implies that the function of MNs cannot be discovered solely by research on their field properties; tests of the kind described in O2F (sect. 8.1.3), and promoted by Lingnau and Caramazza, are also needed. We agree, therefore, that temporary and permanent lesion studies are an important component of research on MN function; indeed, we presented some evidence from such studies in the target article. However, such studies are vulnerable to what we termed the 'localization problem': not all neurons in 'classical' MN areas are mirror neurons and thus lesion studies may over-estimate the involvement of MNs in the cognitive function of interest.

Oosterhof et al. pointed out that MVPA, while not a causal technique, can help to overcome the localization problem. This method, unlike standard fMRI approaches, allows measurement of spatially distributed responses to *specific* observed and executed actions. Oosterhof et al. provided an interesting description of how this technique could be used to test further the predictions of the associative account, and we agree that MVPA may also prove a promising additional tool to investigate the functional properties of MNs.

Bonaiuto argued that computational models can provide a convincing demonstration of the concepts underlying the associative account. We agree, and have therefore previously presented such a model (Cooper, Catmur, & Heyes, 2013; Cooper, Cook, Dickinson, & Heyes, 2013). As Bonaiuto suggested, this model requires appropriate input representations of actions. It demonstrates how sensorimotor associative learning can modulate the development of mirror properties (Cooper, Cook, et al., 2013) and how mirror networks operate, once developed, to produce imitative behavior (Cooper, Catmur, & Heyes, 2013). Other computational models which

use domain-general learning processes and demonstrate the development of MN-like properties in the domain of speech processing were described by **Holt and Lotto, and Saygin and Dick. Giese** also summarized how insights from visual pattern recognition may be applied to computational modeling of MNs. Thus we consider computational modeling to be important not only for testing accounts of how MNs develop, but also for providing novel predictions regarding the functional properties of mature mirror systems.

R8 Conclusion

We are grateful to all of our commentators, combatants and supporters alike, for studying O2F and contributing their insights and opinions to the debate. Given our focus on tests and testability – both in comparing the associative account with genetic and hybrid models of the origin of MNs, and in future research on the function of MNs – we found the methodological and empirically-grounded commentaries especially valuable, but they were all both thoughtful and provocative. The contributions from **Rizzolatti** and other members of the Parma group were especially welcome. We disagree with the Parma group on a number of points, but we have no doubt that their shining discovery, MNs, will intrigue scientists and philosophers for many years to come.

References

Arnstein, D., Cui, F., Keysers, C., Maurits, N. M., & Gazzola, V. (2011). μ-suppression during action observation and execution correlates with BOLD in dorsal premotor, inferior parietal, and SI cortices. *Journal of Neuroscience,31*(40), 14243-9. doi: 10.1523/JNEUROSCI.0963-11.2011.

Ayres, J. J. B., Haddad, C., & Albert, M. (1987). One-trial excitatory backward conditioning as assessed by conditioned suppression of licking in rats: concurrent observations of lick suppression and defensive behaviors. *Animal Learning and Behavior, 15*, 212-217.

Bird, G., Leighton, J., Press, C., & Heyes, C. (2007). Intact automatic imitation of human and robot actions in autism spectrum disorders. *Proceedings of the Royal Society B: Biological Sciences, 274*(1628), 3027-31.

Cook, R. (2012). The ontogenetic origins of mirror neurons: evidence from 'tool-use' and 'audiovisual' mirror neurons. *Biology Letters, 8*(5), 856-9. doi: 10.1098/rsbl.2012.0192.

Cook, R., Dickinson, A., & Heyes, C. (2012). Contextual modulation of mirror and countermirror sensorimotor associations. *Journal of Experimental Psychology: General, 141*(4), 774-787.

Cooper, R. P., Catmur, C., & Heyes, C. (2013). Are automatic imitation and spatial compatibility mediated by different processes? *Cognitive Science, 37*(4), 605-630.

Cooper, R. P., Cook, R., Dickinson, A., & Heyes, C. M. (2013). Associative (not Hebbian) learning and the mirror neuron system. *Neuroscience Letters, 540*, 28-36.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy,* LXXII, 741-765.

Dick, F., Lee, H.L., Nusbaum, H., & Price, C.J. (2011). Auditory-motor expertise alters “speech selectivity” in professional musicians and actors. *Cerebral Cortex*, 21(4):938-48.

Dushanova, J., & Donoghue, J. (2010). Neurons in primary motor cortex engaged during action observation. *European Journal of Neuroscience, 31*(2), 386-398.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 229-40.

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action-effect learning. *Psychological Research, 68*(2-3), 138-54.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17*(8), 1703-1714.

Ferrari, P. F., Visalberghi, E., Paukner, A., Fogassi, L., Ruggiero, A., & Suomi, S. J. (2006). Neonatal imitation in rhesus macaques. *PLoS Biol*ogy, *4*(9), e302.

Garcia, J., Kimeldorf, D. J., & Koelling, R. A. (1955). Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science, 122*, 157-8.

Gemberling, G. A., & Domjan, M. (1982). Selective association in one-day old rats: Taste-toxicosis and texture-toxicosis aversion learning. *Journal of Comparative and Physiological Psychology*, *96*, 105-113.

Godfrey-Smith, P. (1994). A modern history theory of functions. *Noûs,* 28, 344-362.

Hari, R., & Salmelin, R. (1997). Human cortical oscillations: a neuromagnetic view through the skull. *Trends in Neurosciences, 20*(1), 44-9.

Heyes, C. M. (2003). Four routes of cognitive evolution. *Psychological Review*, *110*, 713-727.

Heyes, C. M. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioural Reviews, 34,* 575-583.

Heyes, C. M. (2011). Automatic imitation. *Psychological Bulletin, 137*(3), 463-483.

Heyes, C. M. (2013). What can imitation do for cooperation? In: Sterelny, K., Joyce, R., Calcott, B., & Fraser, B. (Eds.), *Cooperation and its evolution*. Cambridge, Mass.: MIT Press.

Heyes, C. M. (under review). Tinbergen on mirror neurons. *Philosophical Transactions of the Royal Society Series B: Biological Sciences.*

Huang, C. T., Heyes, C., & Charman, T. (2002). Infants' behavioral reenactment of "failed attempts": exploring the roles of emulation learning, stimulus enhancement, and understanding of intentions. *Developmental Psychology, 38*(5), 840-55.

Jones, S. S. (1996). Imitation or exploration? Young infants' matching of adults' oral gestures. *Child Development, 67*(5), 1952-69.

Jones, S. S. (2006). Exploration or imitation? The effect of music on 4-week-old infants' tongue protrusions. *Infant Behavior and Development, 29*(1), 126-30.

Keysers, C., Kohler, E., Umiltà, M. A., Nanetti, L., Fogassi, L., & Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research, 153*(4), 628-36.

Kraskov, A., Dancause, N., Quallo, M. M., Shepherd, S., & Lemon, R.N. (2009). Corticospinal neurons in macaque ventral premotor cortex with mirror properties: a potential mechanism for action suppression? *Neuron, 64*(6):922-30. doi: 10.1016/j.neuron.2009.12.010.

Kunde, W. (2001). Response-effect compatibility in manual choice reaction tasks. *Journal of Experimental Psychology: Human Perception and Performance, 27*(2), 387-94.

Lepage, J. F., Saint-Amour, D., & Théoret, H. (2008). EEG and neuronavigated single-pulse TMS in the study of the observation/execution matching system: are both techniques measuring the same process? *Journal of Neuroscience Methods, 175*(1), 17-24. doi: 10.1016/j.jneumeth.2008.07.021.

Liepelt, R., von Cramon, D. Y., & Brass, M. (2008). What is matched in direct matching? Intention attribution modulates motor priming. *Journal of Experimental Psychology: Human Perception and Performance, 34*(3), 578-91.

Linquist, S., Machery, E., Griffiths, P. E. & Stotz, K. (2011) Exploring the folkbiological conception of human nature. *Philosophical Transactions of the Royal Society Series B: Biological Sciences*, *366*, 444–453. doi:10.1098/rstb.2010.0224.

Mackintosh, N. J. (1973). Stimulus selection: Learning to ignore stimuli that predict no change in reinforcement. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Constraints on learning* (pp. 75-100). London: Academic Press.

Mahoney, W. J., & Ayres, J. J. B. (1976). One-trial simultaneous and backward fear conditioning as reflected in conditioned suppression of licking in rats. *Animal Learning and Behavior, 4*(4), 357-362.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology, 20*(8), 750-756.

Nagy, E., Pilling, K., Orvos, H., & Molnar, P. (2012). Imitation of Tongue Protrusion in Human Neonates: Specificity of the Response in a Large Sample. *Developmental Psychology,* in press.

Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2011). Action observation circuits in the macaque monkey cortex. *Journal of Neuroscience, 31*(10), 3743-56. doi: 10.1523/JNEUROSCI.4803-10.2011.

Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2005). Observing others: multiple action representation in the frontal lobe. *Science 310*(5746), 332-6.

Paukner, A., Ferrari, P. F., & Suomi, S. J. (2011). Delayed imitation of lipsmacking gestures by infant rhesus macaques (Macaca mulatta). *PLoS One 6*(12):e28848. doi: 10.1371/journal.pone.0028848.

Pearce, J. M. (2008). Animal learning and conditioning: an introduction. Third edition. Hove and New York: Psychology Press.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*(1), 176-80.

Prather, J.F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature, 451*(7176), 305-310.

Press, C. (2011). Action observation and robotic agents: Learning and anthropomorphism. *Neuroscience and Biobehavioural Reviews*, *35,* 1410-1418.

Press, C., Gillmeister, H., & Heyes C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proceedings of the Royal Society B: Biological Sciences, 274*(1625), 2509-14.

Press, C., Heyes, C., & Kilner, J. M. (2011). Learning to understand others' actions. *Biology Letters, 7*(3), 457-60. doi: 10.1098/rsbl.2010.0850.

Press, C., Richardson, D., & Bird, G. (2010). Intact imitation of emotional facial actions in autism spectrum conditions. *Neuropsychologia*, *48*(11), 3291-7. doi: 10.1016/j.neuropsychologia.2010.07.012.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: II. Current research and theory*, pp. 64-99. New York, NY: Appleton-Century-Crofts.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience, 11*(4), 264-74. doi: 10.1038/nrn2805.

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience*, *23*, 473-500.

Southgate, V,. & Hamilton, A. F. (2008). Unbroken mirrors: challenging a theory of Autism. *Trends in Cognitive Sciences, 12*(6), 225-9. doi: 10.1016/j.tics.2008.03.005.

Tiedens, L. Z., & Fragale, A. R. (2003). Power moves: Complementarity in dominant and submissive nonverbal behavior. *Journal of Personality and Social Psychology, 84*(3), 558-568.

Tkach, D., Reimer, J., & Hatsopoulos, N. G. (2007). Congruent activity during action and action observation in motor cortex. *Journal of Neuroscience, 27*(48), 13241-13250.

Vigneswaran, G., Philipp, R., Lemon, R. N., & Kraskov, A. (2013). M1 corticospinal mirror neurons and their role in movement suppression during action observation. *Current Biology, 23*(3), 236-43. doi: 10.1016/j.cub.2012.12.006.

Wiggett, A. J., Hudson, M., Tipper, S. P., & Downing, P. E. (2011). Learning associations between action and perception: effects of incompatible training on body part and spatial priming. *Brain and Cognition, 76*(1), 87-96. doi: 10.1016/j.bandc.2011.02.014.

[1] Contrary to what some commentators suggested, we have demonstrated this not only behaviorally but with all of the neurophysiological measures commonly used to detect MNs in humans (**Bertenthal**), and with measures that demonstrate action-specificity (**Oosterhof, Wiggett, & Cross**).

[2] We are very grateful to Nick Shea for suggesting this format.